

Historical Analysis of James River Water Quality Data

Harry Wang*, Fei Ye, and Dave Forrest

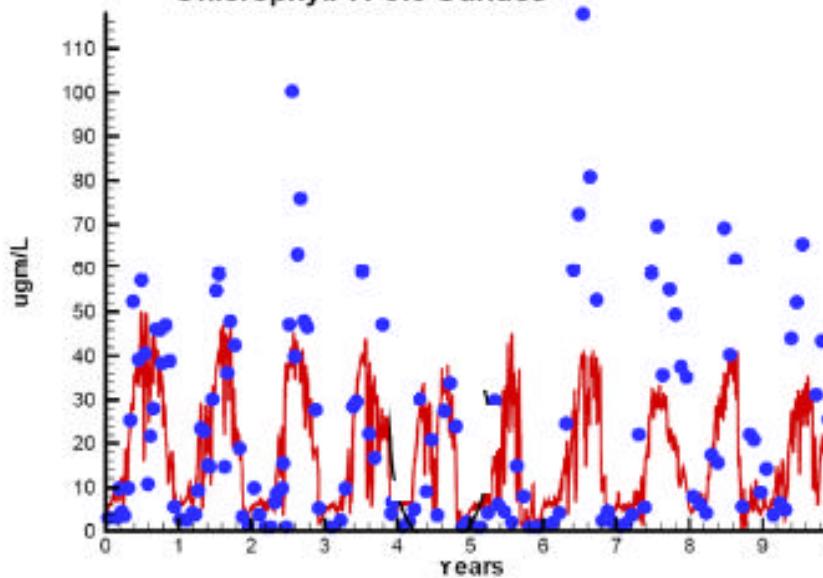
Virginia Institute of Marine Sciences

Gloucester Point, VA 23062

*email:hvwwang@vims.edu

James river SAP meeting on 04/26/2013

56920 Grid (Run 417)
Chlorophyll TF5.5 Surface



56920 Grid (Run 417)
Chlorophyll LE5.3 Surface

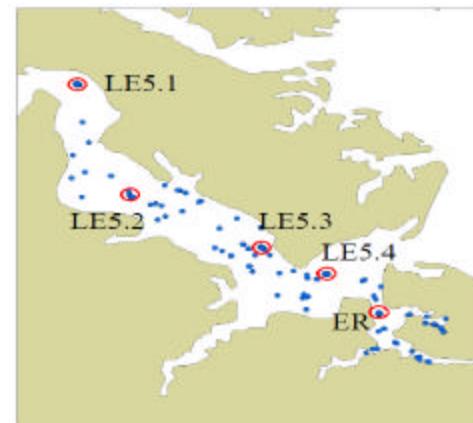
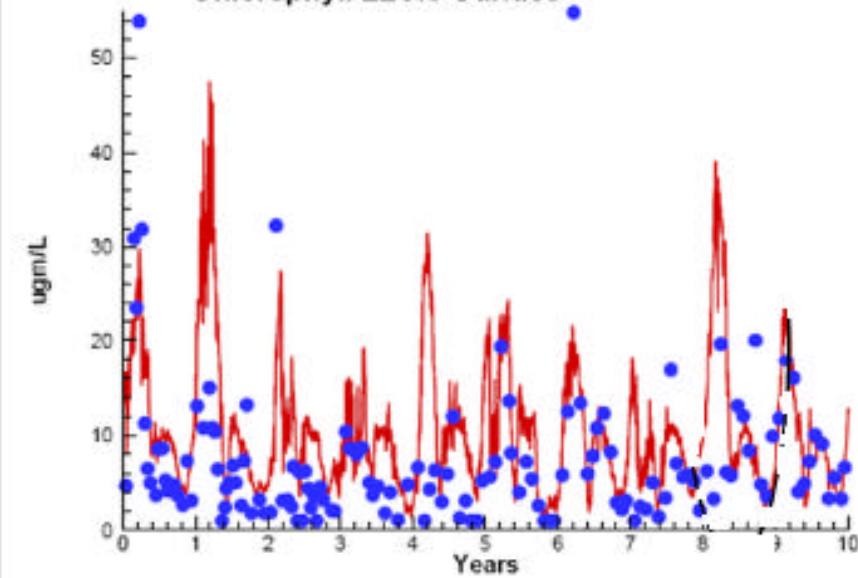


Figure: The water quality model results from Chesapeake Bay 50,000 cells version

Strategy: Isolate temporal and spatial variability

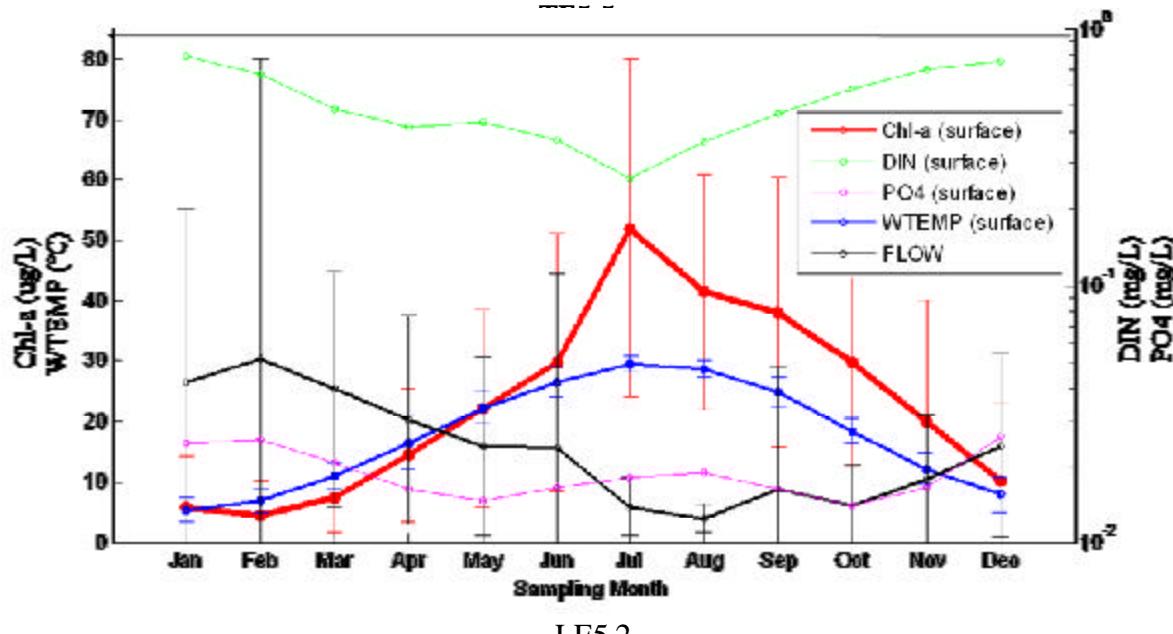
Ensemble average

$\langle f \rangle$ = ensemble average monthly data over 20 years (and 10 years) periods from 1991-2010.

- $U = \langle U \rangle + U'$
- $C = \langle C \rangle + C'$
- $T = \langle T \rangle + T'$

Cluster analysis

- A cluster analysis was performed on the James River data based on the dissimilarity of the sampled Chl-a concentration among the stations.
- The dissimilarity is measured by Euclidean distance between station groups, in which each station is deemed as a vector with N components (N is the sample size).



TF5.5

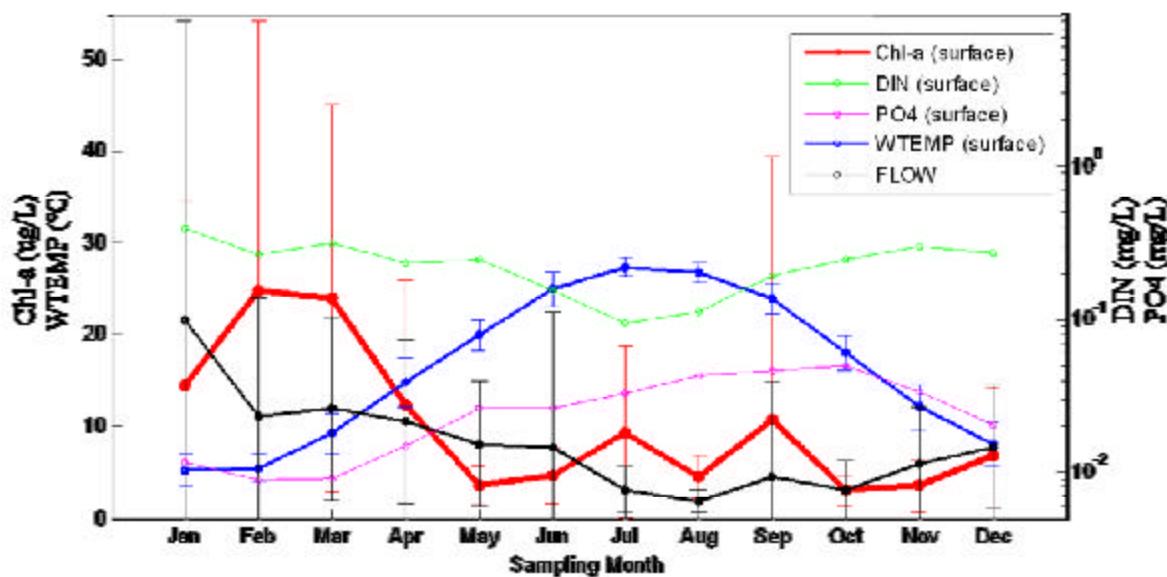


Figure 37: Seasonal cycles in water quality data at the locations of the Chl-a maximums. Top panel: TF5.5 July maximum; Bottom panel LE5.2 February maximum. Note the high correlation of Chl-a with temperature in the upper panel, while the correlation is opposite in the lower panel. (The flow are normalized and scaled with max. Chl-a).

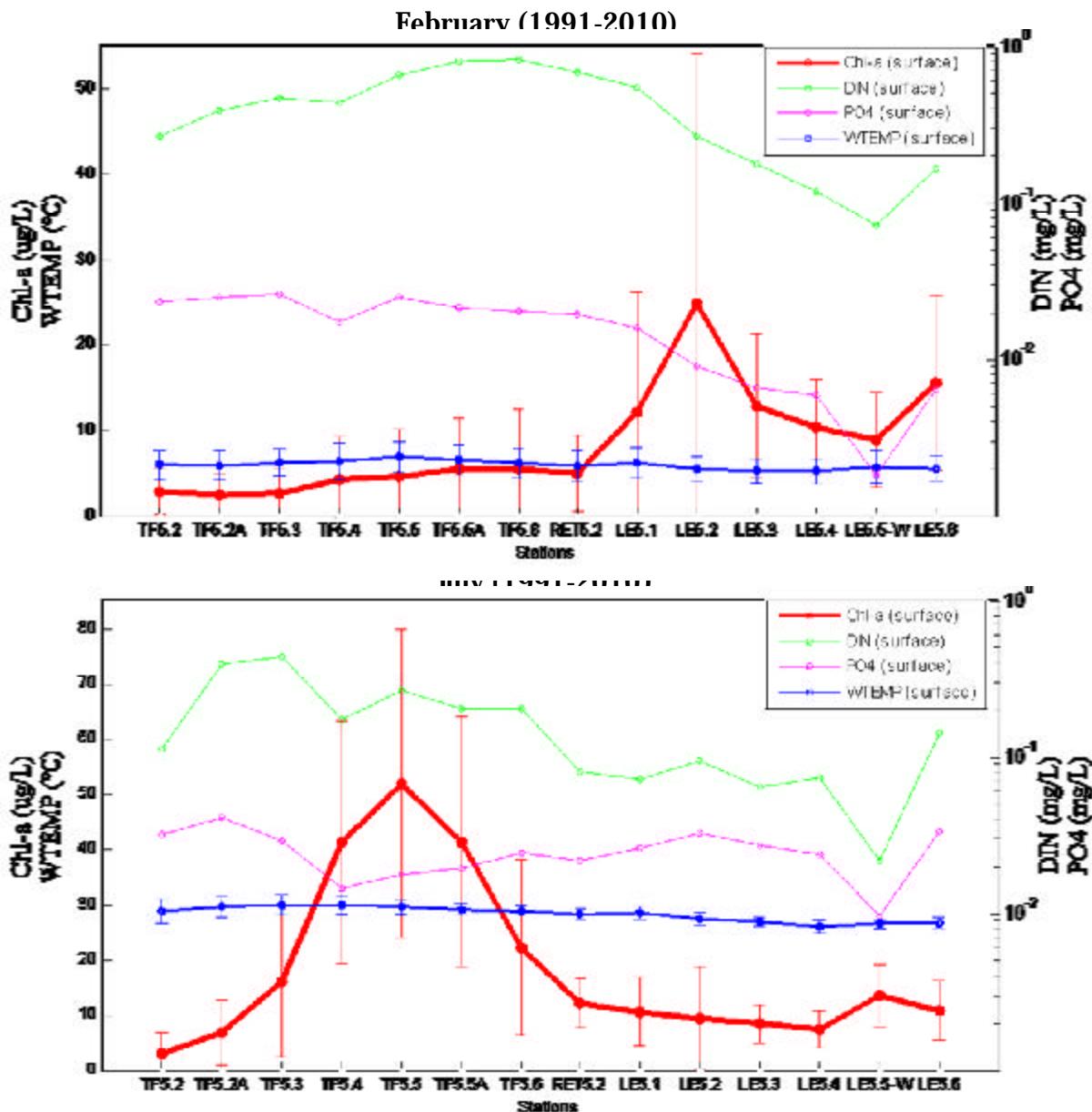
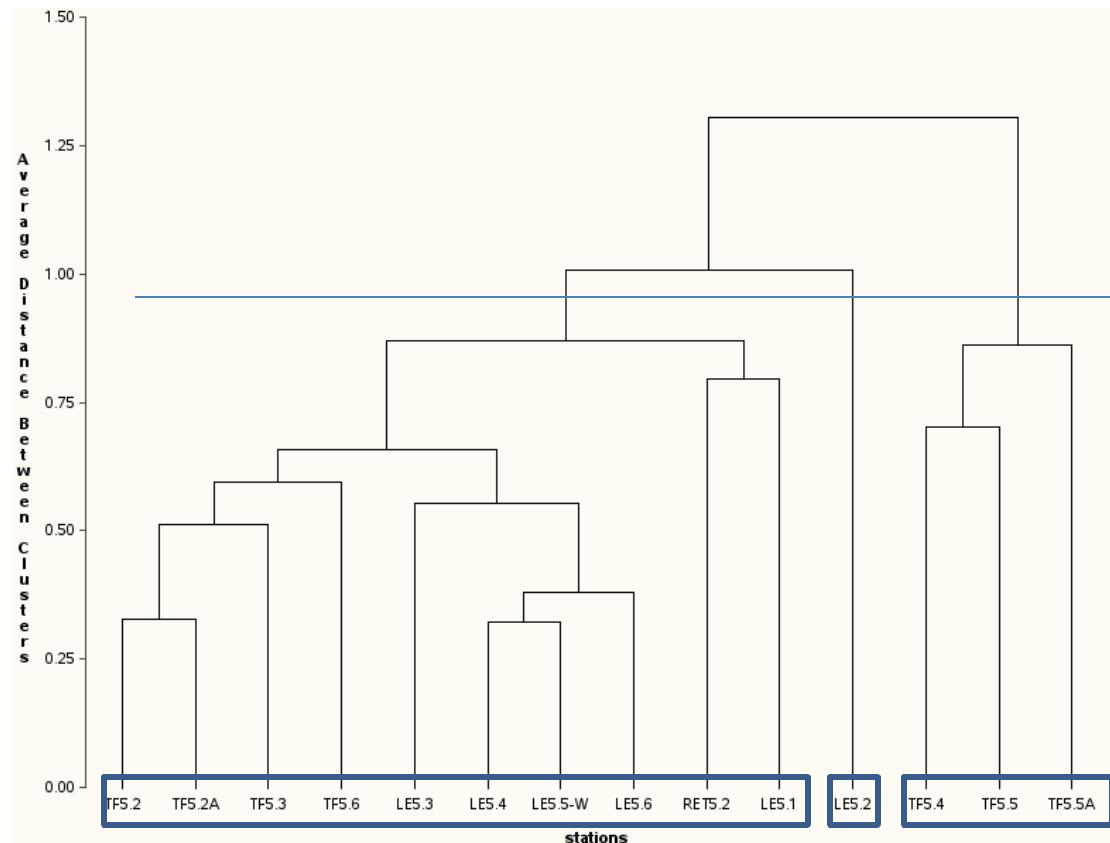


Figure 36: James River spatial variability of monthly averaged Chl-a, DIN, PO₄, and water temperature for February (top panel) and July (bottom panel), 1991-2010. Note the February Chl-a maximum is located at station LE5.2, and the July Chl-a maximum at station TF5.5.

James River: Cluster Analysis (1991~2010)

Cluster History					
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square
13	LE5.4	LE5.5-W	2	0.0079	0.992
12	TF5.2	TF5.2A	2	0.0082	0.984
11	CL13	LE5.6	3	0.0121	0.972
10	CL12	TF5.3	3	0.0242	0.948
9	LE5.3	CL11	4	0.0303	0.917
8	CL10	TF5.6	4	0.0325	0.885
7	CL8	CL9	8	0.0753	0.809
6	TF5.4	TF5.5	2	0.0378	0.772
5	RET5.2	LE5.1	2	0.0488	0.723
4	CL6	TF5.5A	3	0.0633	0.66
3	CL7	CL5	10	0.1093	0.55
2	CL3	LE5.2	11	0.1102	0.44
1	CL2	CL4	14	0.44	0



14 stations;
223 samples per station

James River: Cluster Analysis (1991~2010)

Obs	stations	TF5.2 3	TF5.2A	TF5.3 0	TF5.4 -	TF5.5 -	TF5.5A	TF5.6 -	RET5.2	LE5.1 -	LE5.2 -	LE5.3 -	LE5.4 1	LE5.5-W	LE5.6 -
1	TF5.2 3	0
2	TF5.2A	83.246	0
3	TF5.3 0	144.965	114.049	0
4	TF5.4 -	313.172	287.768	251.687	0
5	TF5.5 -	332.97	308.278	258.821	175.929	0
6	TF5.5A	309.409	292.176	273.307	244.375	193.408	0
7	TF5.6 -	150.743	150.074	151.856	262.189	262.216	238.52	0
8	RET5.2	175.399	197.916	210.35	334.112	345.157	316.381	196.752	0
9	LE5.1 -	210.148	236.752	252.694	374.14	387.005	367.814	253.398	203.122	0
10	LE5.2 -	237.772	258.065	278.46	411.838	418.419	387.619	282.25	289.711	260.824	0
11	LE5.3 -	162.084	178.535	208.839	344.457	358.593	345.418	202.662	228.026	253.72	248.282	0	.	.	.
12	LE5.4 1	111.272	138.776	173.493	331.492	348.698	335.682	174.862	190.391	224.981	225.34	129.67	0	.	.
13	LE5.5-W	125.502	145.571	168.386	324.851	344.525	342.572	183.787	204.993	235.991	243.217	152.728	81.5799	0	.
14	LE5.6 -	137.4	156.467	187.535	332.615	349.48	336.892	187.907	215.254	240.272	232.961	139.03	87.5325	104.622	0

This is the distance matrix of the cluster analysis in the previous slide. Each entry is the Euclidean distance between a pair of 2 stations.

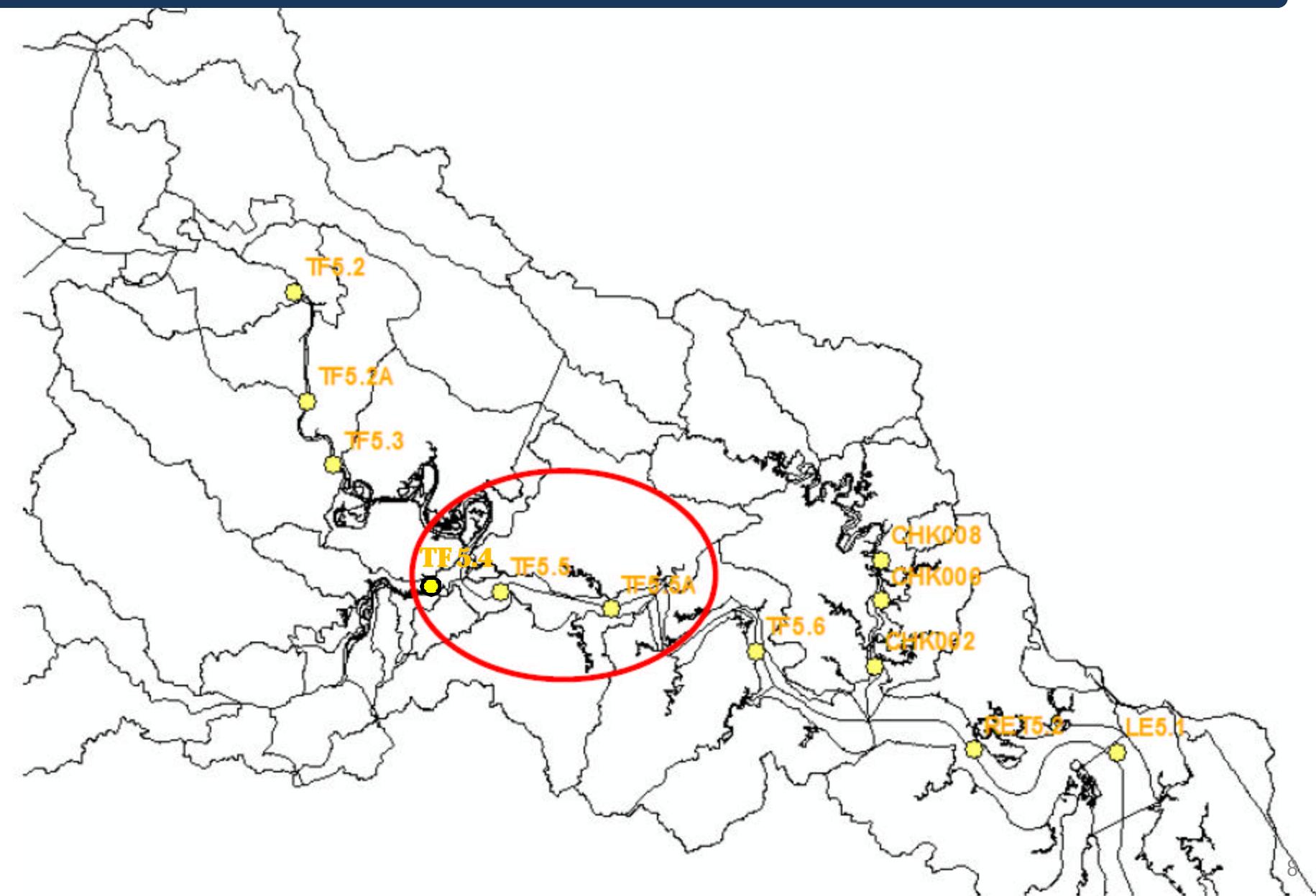
Root-Mean-Square Distance Between Observations = 255, which can be calculated by:

$$\sqrt{\sum D_{i,j}^2}$$

, where D are the distances in the above table.

The “tree plot” in the previous slide shows the “average distance between clusters”, which is the **Euclidean distance between clusters scaled (divided) by the Root-Mean-Square Distance Between Observations** (255 in this case)

Station group (mid-stream, location of summer peak: TF5.4 TF5.5 TF5.5A)



Station group (mid-stream, location of summer peak: TF5.4 TF5.5 TF5.5A)

Correlation Matrix, Monthly ensemble average, 3 periods, 12 data points (12 months) per period per parameter

	Pearson Correlation Coefficients, N = 12								
	time periods	Prob > r under H0: Rho=0							
		logChl-a	FLOW	WTEMP	SALINITY	TP	SECCHI	TSS	TN
logChl-a	1991~2010	1							
	1991~2000		1						
	2001~2010			1					
FLOW	1991~2010	-0.91907	1						
	1991~2000	-0.92008		1					
	2001~2010	-0.68059			1				
WTEMP	1991~2010	0.93578	-0.74733	1					
	1991~2000	0.89374	-0.75288		1				
	2001~2010	0.93365	-0.506			1			
SALINITY	1991~2010	0.18864	-0.47444	-0.1219	1				
	1991~2000	0.35355	-0.55445	0.03863		1			
	2001~2010	0.02549	-0.28376	-0.14344			1		
TP	1991~2010	0.78934	-0.77547	0.73084	0.16487	1			
	1991~2000	0.72748	-0.66365	0.71037	0.35407		1		
	2001~2010	0.77014	-0.56879	0.6469	0.16872			1	
SECCHI	1991~2010	-0.01836	0.164	0.11653	-0.39527	-0.49565	1		
	1991~2000	0.24988	-0.25321	0.31109	-0.10882	-0.35562		1	
	2001~2010	-0.32444	0.30087	-0.17227	-0.46214	-0.69381			1
TSS	1991~2010	-0.20634	0.16696	-0.2259	0.09922	0.37278	-0.83025	1	
	1991~2000	0.01189	0.01589	-0.12806	0.31588	0.56118	-0.90651		1
	2001~2010	-0.20443	0.4756	-0.19096	0.04276	0.34262	-0.47294		1
TN	1991~2010	0.34361	-0.58853	0.08927	0.78943	0.55498	-0.6707	0.43062	1
	1991~2000	0.58745	-0.75541	0.32881	0.63008	0.59443	-0.17968	0.30333	
	2001~2010	-0.24082	-0.14105	-0.46498	0.59556	0.30366	-0.66595	0.48335	

Station group (mid-stream, location of summer peak: TF5.4 TF5.5 TF5.5A)

Summary of regression analysis between Chl-a and each parameter

Raw Data

Monthly (ensemble average, 1991~2010)

Statistical Model:

$$\log(CHLA) = a + bx + \epsilon$$

(Linear regression, log transform on Chl-a)

Parameters	R ²	samples
WTEMP	0.52	178
SALINITY	<0.05	167
SECCHI	<0.05	174
TSS	0.06	179
TP	0.15	177
TN	0.18	175
FLOW	0.31	180
FLOW5	0.41	180
FLOW10	0.42	180
FLOW20	0.39	180
FLOW30	0.35	180
FLOW40	0.3	180

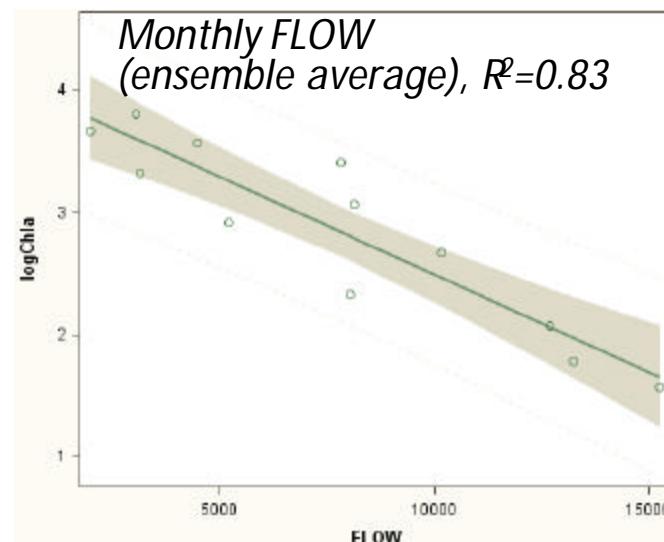
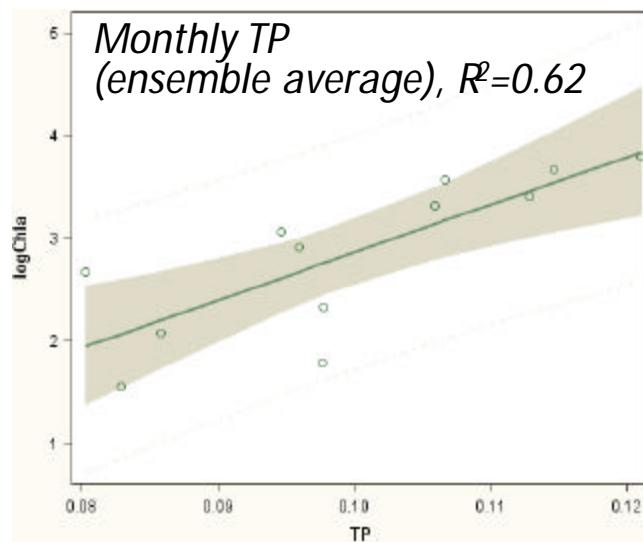
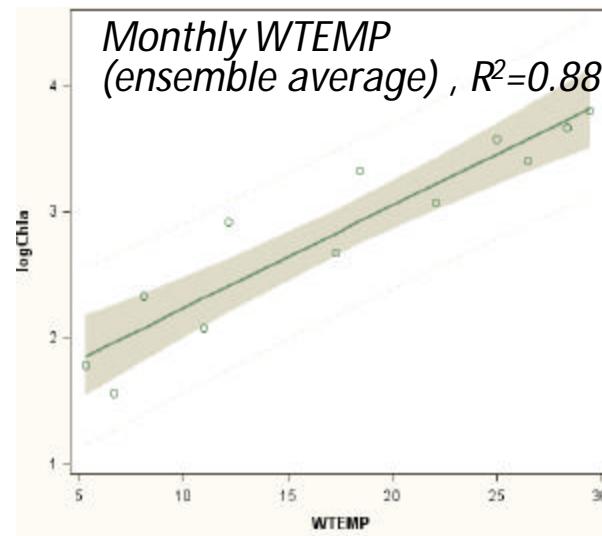
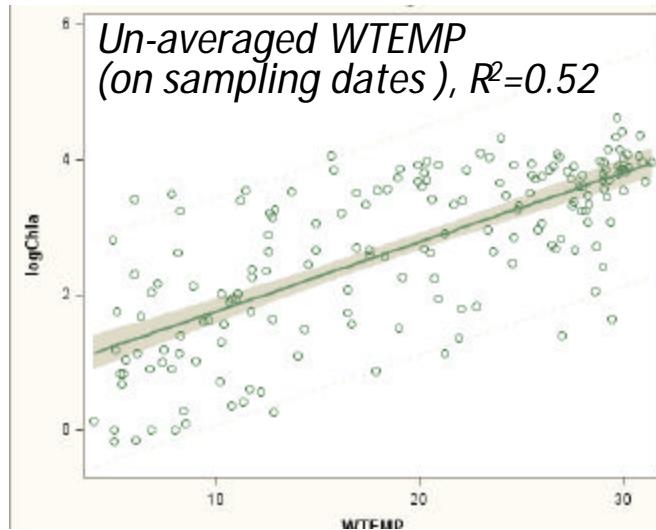
Parameters	R ²
WTEMP	0.88
SALINITY	<0.05
SECCHI	<0.05
TSS	<0.05
TP	0.62
TN	0.12
FLOW	0.83

12 samples per parameter

Selective regression plots
are presented in next slide

Station group (mid-stream, location of summer peak: TF5.4 TF5.5 TF5.5A)

Selective regression plots between Chl-a and each water quality parameter
1991~2010 , (only those with high correlation)



Station group (mid-stream, location of summer peak: TF5.4 TF5.5 TF5.5A)

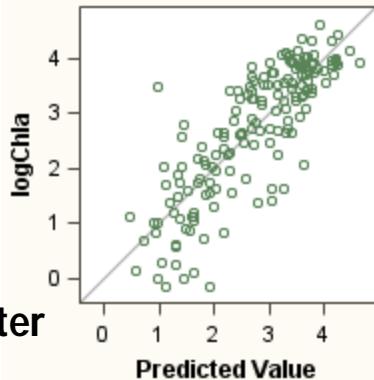
Multiple Regression
Statistical Model:

$$\log(CHLA) = a + b_i x_i + \epsilon$$

(1991~2010)

R²=0.85

167 samples per parameter

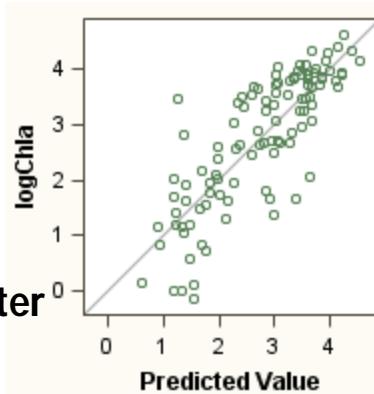


Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	4.08708	0.96166	4.25	<.0001	0
logFlow	1	-0.49385	0.0787	-6.28	<.0001	-0.41136
WTEMP	1	0.06593	0.00806	8.18	<.0001	0.46936
TP	1	-2.2338	2.40024	-0.93	0.3534	-0.05907
SECCHI	1	1.14854	0.56616	2.03	0.0442	0.10596
TSS	1	0.02517	0.0085	2.96	0.0035	0.17625
TN	1	0.38655	0.28598	1.35	0.1784	0.09018

(1991~2000)

R²=0.85

106 samples per parameter

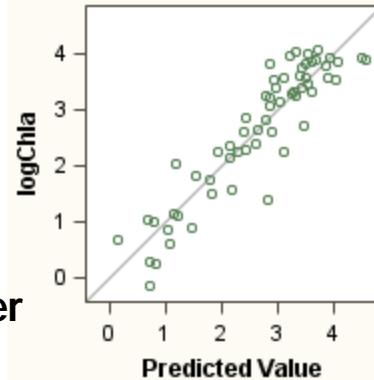


Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	4.03934	1.43219	2.82	0.0058	0
logFlow	1	-0.40036	0.11837	-3.38	0.001	-0.3091
WTEMP	1	0.08441	0.01119	7.54	<.0001	0.59264
TP	1	-6.13343	3.23779	-1.89	0.0611	-0.16099
SECCHI	1	-0.14433	0.80156	-0.18	0.8575	-0.0122
TSS	1	0.02744	0.01195	2.3	0.0238	0.17516
TN	1	0.35747	0.35992	0.99	0.323	0.08773

(2001~2010)

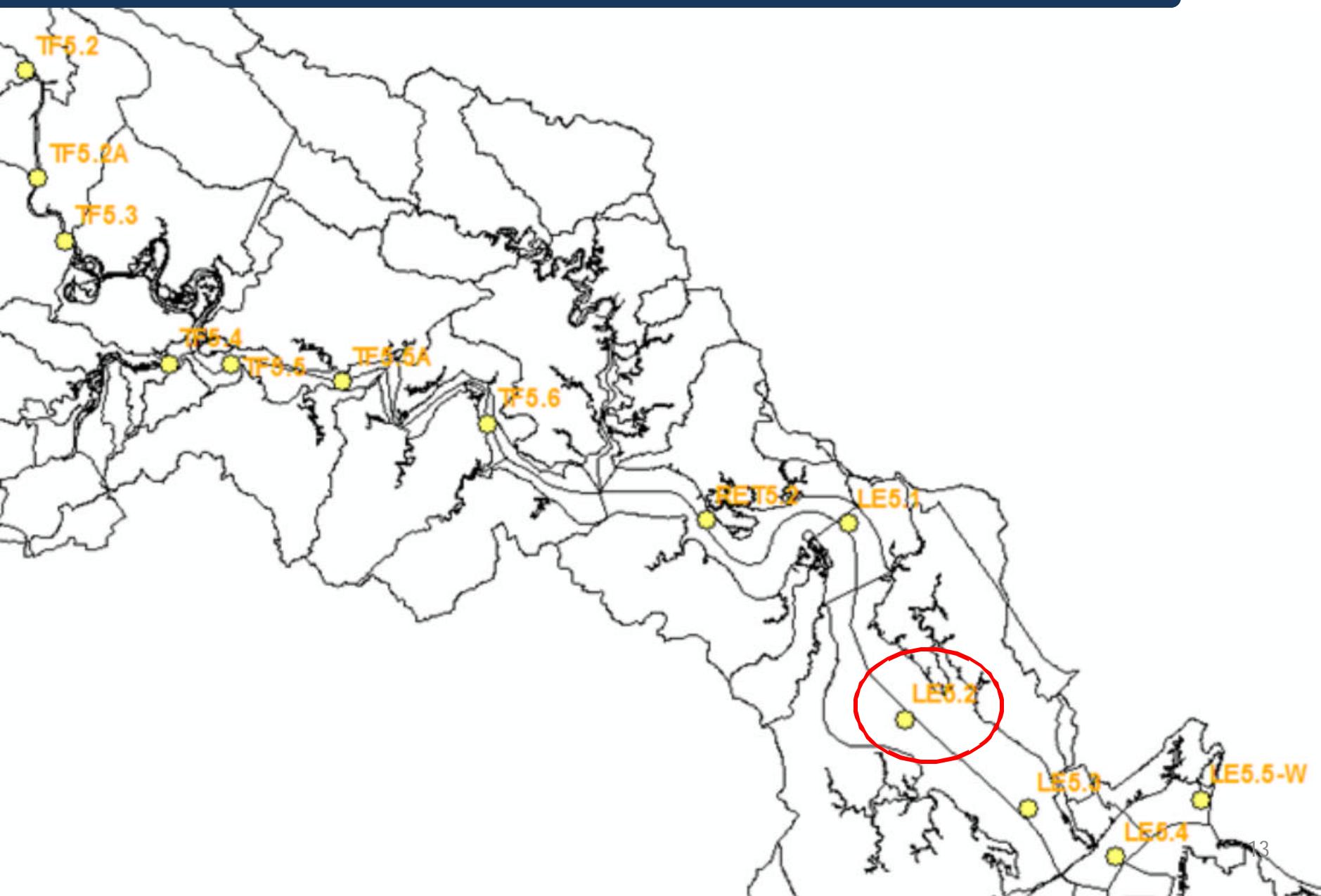
R²=0.85

61 samples per parameter



Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	-0.09815	1.48263	-0.07	0.9475	0
logFlow	1	-0.36097	0.0982	-3.68	0.0005	-0.33621
WTEMP	1	0.05873	0.0102	5.76	<.0001	0.42917
TP	1	4.98124	3.38126	1.47	0.1465	0.11543
SECCHI	1	4.45925	0.87991	5.07	<.0001	0.38683
TSS	1	0.03151	0.00978	3.22	0.0022	0.25188
TN	1	1.35295	0.45993	2.94	0.0048	0.25851

Station group (downstream, location of spring peak: LE5.2)



Station group (downstream, location of spring peak: LE5.2)

Correlation Matrix, Monthly ensemble average, 3 periods, 12 data points (12 months) per period per parameter

	Pearson Correlation Coefficients, N = 12 Prob > r under H0: Rho=0								
	time periods	logChl-a	FLOW	WTEMP	SALINITY	TP	SECCHI	TSS	TN
logChl-a	1991~2010	1							
	1991~2000	1							
	2001~2010	1							
FLOW	1991~2010	0.57436	1						
	1991~2000	0.50055	1						
	2001~2010	0.35689	1						
WTEMP	1991~2010	-0.49558	-0.73337	1					
	1991~2000	-0.369	-0.75697	1					
	2001~2010	-0.54451	-0.5179	1					
SALINITY	1991~2010	-0.54413	-0.63619	0.36606	1				
	1991~2000	-0.54373	-0.70027	0.32033	1				
	2001~2010	-0.32106	-0.47412	0.36794	1				
TP	1991~2010	0.39703	-0.26018	0.41198	-0.10457	1			
	1991~2000	0.68712	0.06711	0.21151	-0.52294	1			
	2001~2010	-0.07507	-0.57795	0.53685	0.65913	1			
SECCHI	1991~2010	-0.51171	-0.34199	-0.06118	0.80554	-0.46531	1		
	1991~2000	-0.52261	-0.4895	-0.00128	0.88091	-0.6237	1		
	2001~2010	-0.30709	-0.0667	-0.18915	0.62692	0.0117	1		
TSS	1991~2010	0.62401	0.47327	-0.21047	-0.93626	0.39533	-0.92429	1	
	1991~2000	0.656	0.48833	-0.18499	-0.91586	0.72747	-0.90496	1	
	2001~2010	0.55445	0.43862	-0.26403	-0.83065	-0.29629	-0.86429	1	
TN	1991~2010	0.82245	0.74822	-0.77347	-0.6585	0.13094	-0.39742	0.62175	1
	1991~2000	0.77461	0.82053	-0.69169	-0.77977	0.48575	-0.54196	0.72082	1
	2001~2010	0.55362	0.41797	-0.70632	-0.21351	-0.20092	0.0661	0.26494	1

Station group (downstream, location of spring peak: LE5.2)

Summary of regression analysis between Chl-a and each water quality parameter

Raw Data

Monthly (ensemble average, 1991~2010)

Statistical Model:

$$\log(CHLA) = a + bx + \epsilon$$

(Linear regression, log transform on Chl-a)

Parameters	R ²	Samples
WTEMP	<0.05	179
SALINITY	<0.05	179
SECCHI	<0.05	177
TSS	<0.05	178
TP	0.05	173
TN	<0.05	171
FLOW	<0.05	179
FLOW5	<0.05	179
FLOW10	<0.05	179
FLOW20	<0.05	179
FLOW30	<0.05	179
FLOW40	<0.05	179

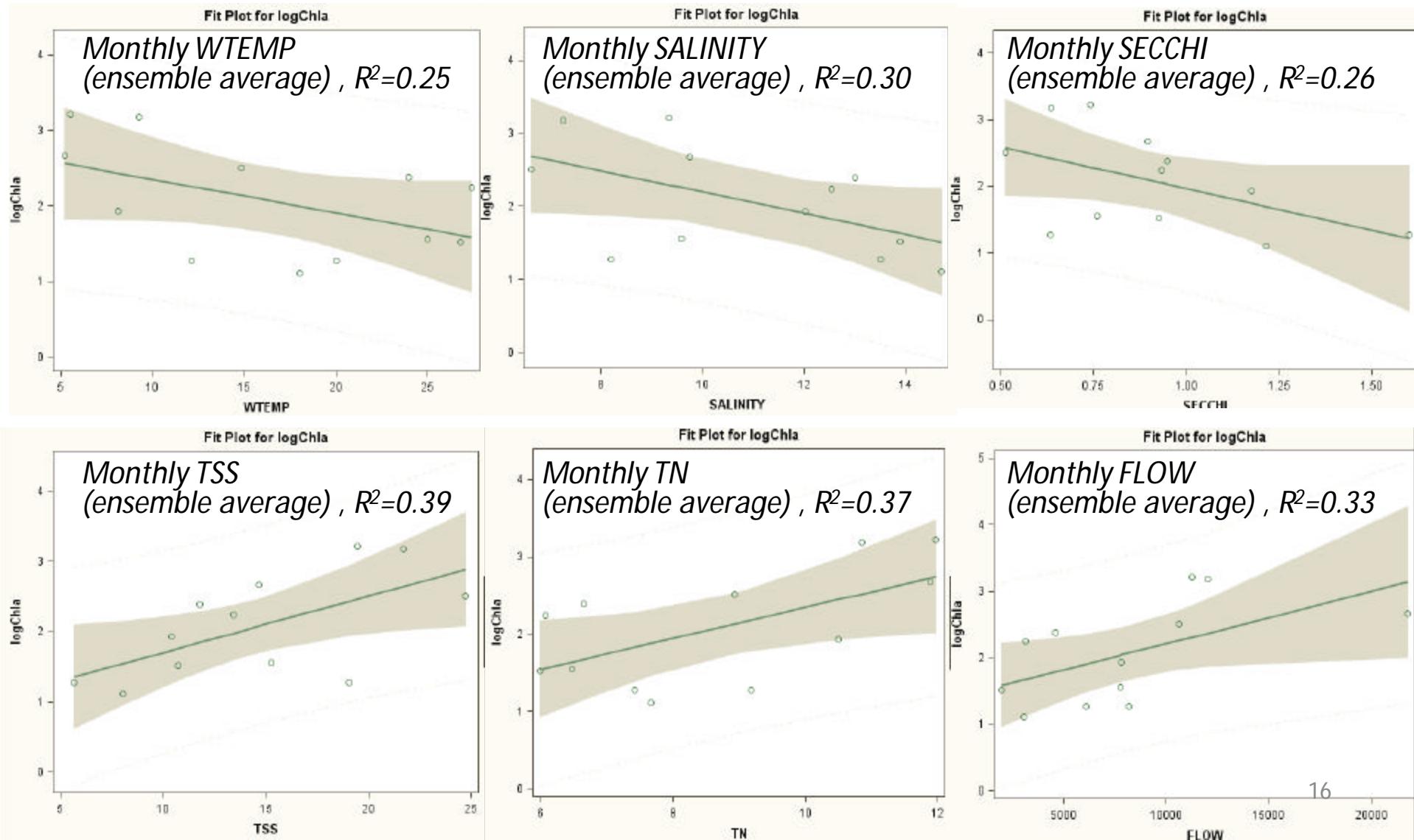
Parameters	R ²
WTEMP	0.25
SALINITY	0.30
SECCHI	0.26
TSS	0.39
TP	0.16
TN	0.68
FLOW	0.33

12 samples per parameter

Selective regression plots
are presented in next slide

Station group (downstream, location of spring peak: LE5.2)

Selective regression plots between Chl-a and each water quality parameter
1991~2010



Station group (downstream, location of spring peak: LE5.2)

Summary of regression analysis between Chl-a and each water quality parameter, monthly ensemble average

Statistical Model:

$$\log(CHLA) = a + bx + \epsilon$$

(Linear regression, log transform on Chl-a)

Parameters	R ² (1991~2010)	R ² (1991~2000)	R ² (2001~2010)
WTEMP	0.25	0.14	0.30
SALINITY	0.30	0.30	0.10
SECCHI	0.26	0.27	0.09
TSS	0.39	0.43	0.31
TP	0.16	0.47	<0.05
TN	0.37	0.60	0.31
FLOW	0.33	0.25	0.13

12 samples per period per parameter

Comments: similar among different periods

Station group (downstream, location of spring peak: LE5.2)

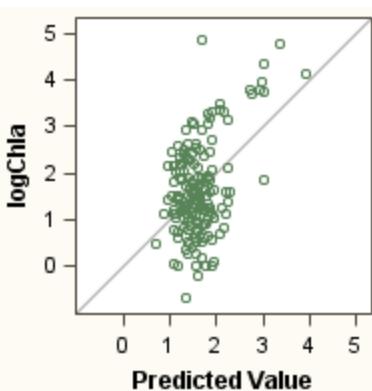
Multiple Regression Statistical Model:

$$\log(CHLA) = a + b_i x_i + \epsilon$$

(1991~2010)

$R^2=0.22$

165 samples per parameter

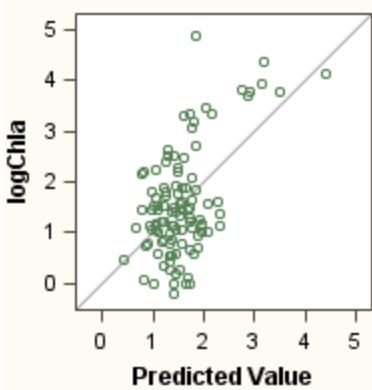


Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	0.88516	1.86268	0.48	0.6353	0
logFlow	1	-0.00794	0.16507	-0.05	0.9617	-0.00763
WTEMP	1	-0.00435	0.01278	-0.34	0.7339	-0.034
SALINITY	1	0.06342	0.03355	1.89	0.0606	0.3108
TP	1	-2.38681	4.09392	-0.58	0.5607	-0.06965
SECCHI	1	-0.7709	0.2871	-2.69	0.008	-0.30756
TSS	1	-0.00674	0.0122	-0.55	0.5811	-0.06132
TN	1	1.70564	0.40408	4.22	<.0001	0.50027

(1991~2000)

$R^2=0.32$

107 samples per parameter

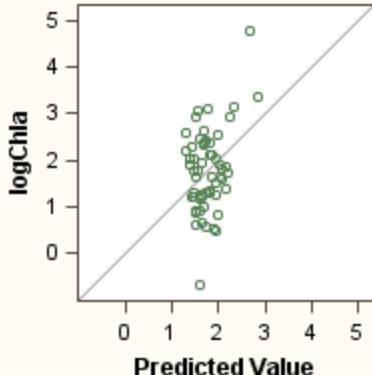


Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	0.10887	2.31887	0.05	0.9626	0
logFlow	1	0.00929	0.20305	0.05	0.9636	0.00797
WTEMP	1	0.00924	0.01727	0.53	0.5939	0.06828
SALINITY	1	0.07634	0.04089	1.87	0.0649	0.33206
TP	1	-2.40412	4.99472	-0.48	0.6313	-0.0737
SECCHI	1	-0.84129	0.3347	-2.51	0.0136	-0.33591
TSS	1	-0.01397	0.01404	-0.99	0.3223	-0.13149
TN	1	2.21021	0.52036	4.25	<.0001	0.64839

(2001~2010)

$R^2=0.13$

58 samples per parameter



Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	0.63299	3.28644	0.19	0.848	0
logflow	1	0.0589	0.30274	0.19	0.8465	0.07105
WTEMP	1	-0.01058	0.02006	-0.53	0.6003	-0.09465
SALINITY	1	0.04352	0.06216	0.7	0.4872	0.26759
TP	1	-2.61814	8.80402	-0.3	0.7674	-0.05484
SECCHI	1	-0.33491	0.61766	-0.54	0.5901	-0.13501
TSS	1	0.0303	0.02581	1.17	0.2459	0.23269
TN	1	0.68888	0.67023	1.03	0.309	0.20332

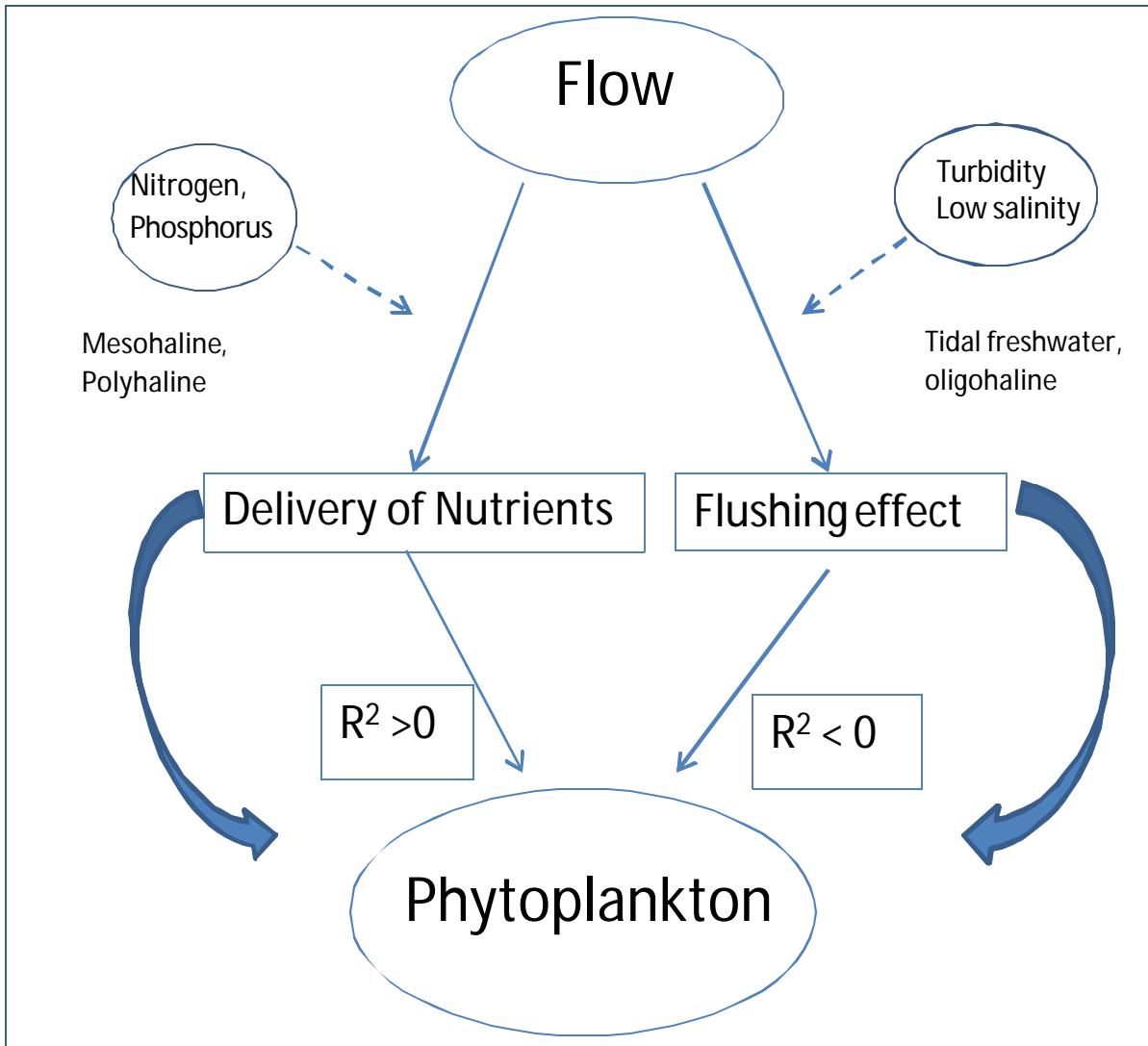
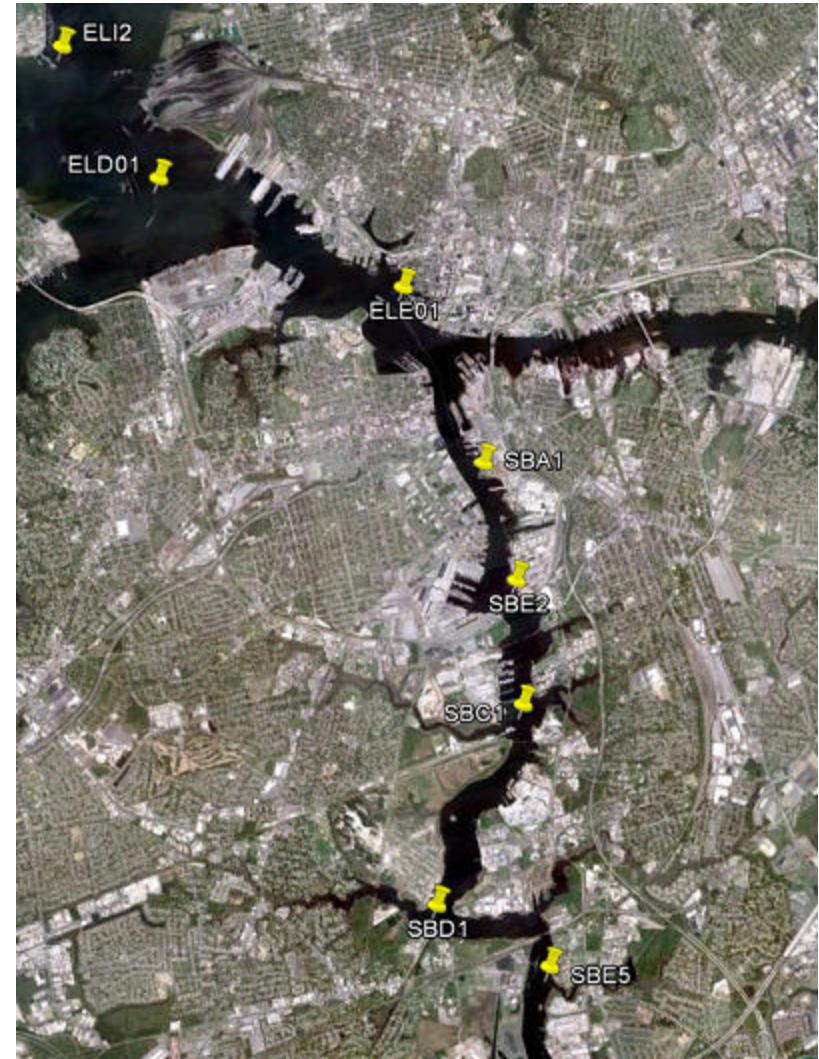
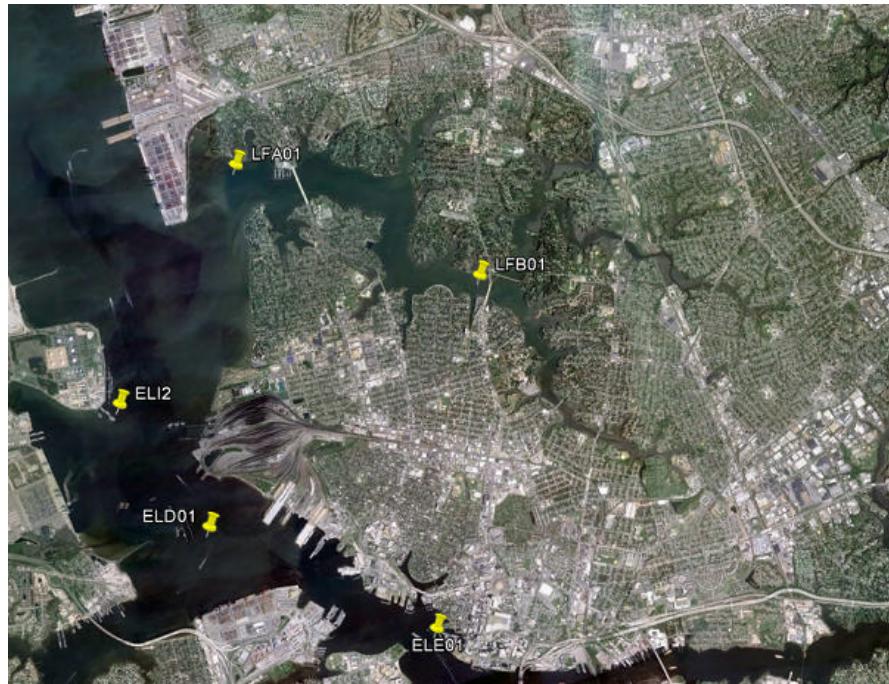


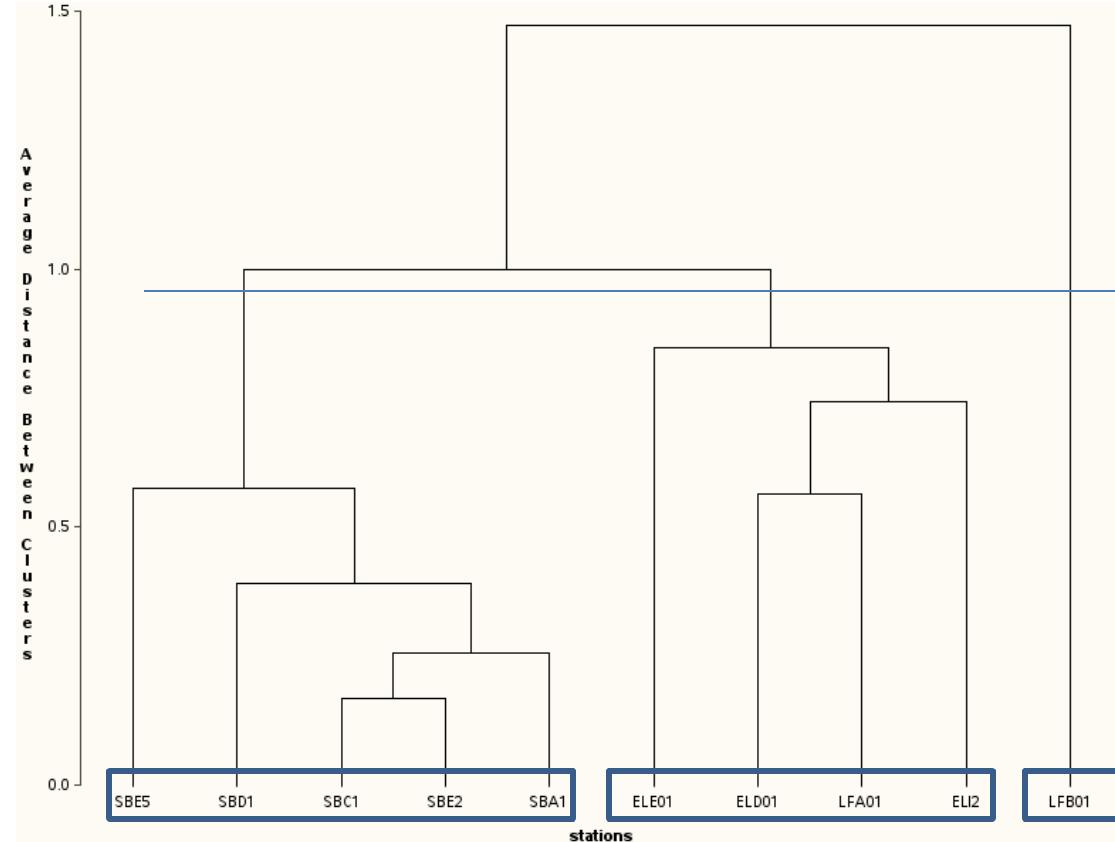
Figure 39: The flow has confounding effects on the phytoplankton. For the tidal freshwater, zone, the flushing effect accompanied by turbidity, and low salinity can dominate (right). For the mesohaline zone, the delivery of nutrients becomes dominate.

Analysis of Elizabeth River water quality data



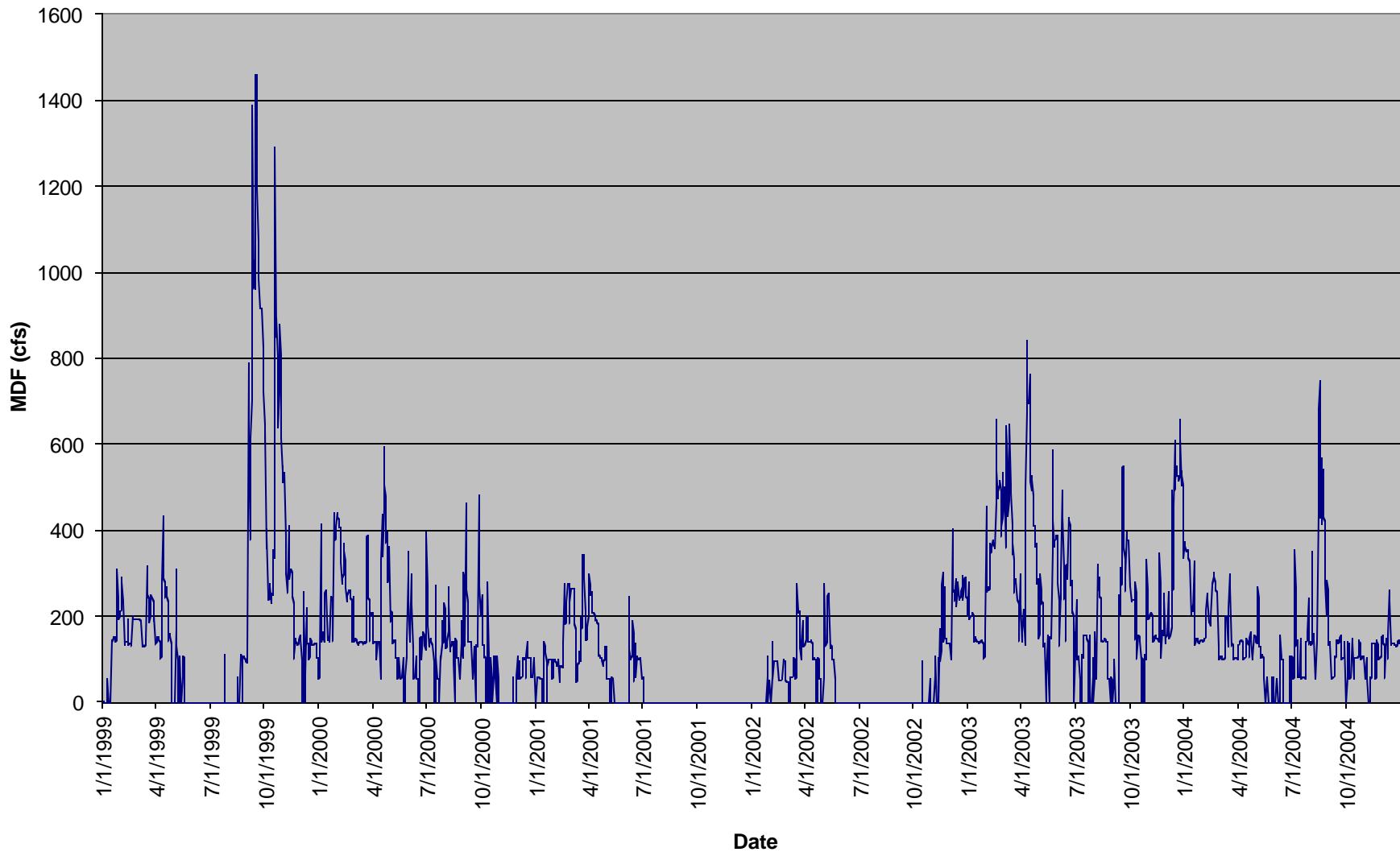
Elizabeth River: Cluster Analysis (1991~2010)

Cluster History					
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square
9	SBC1	SBE2	2	0.0031	0.997
8	CL9	SBA1	3	0.0086	0.988
7	SBD1	CL8	4	0.0224	0.966
6	ELD01	LFA01	2	0.0354	0.93
5	SBE5	CL7	5	0.0518	0.879
4	CL6	ELI2	3	0.0698	0.809
3	ELE01	CL4	4	0.0932	0.716
2	CL5	CL3	9	0.3459	0.37
1	CL2	LFB01	10	0.3698	0

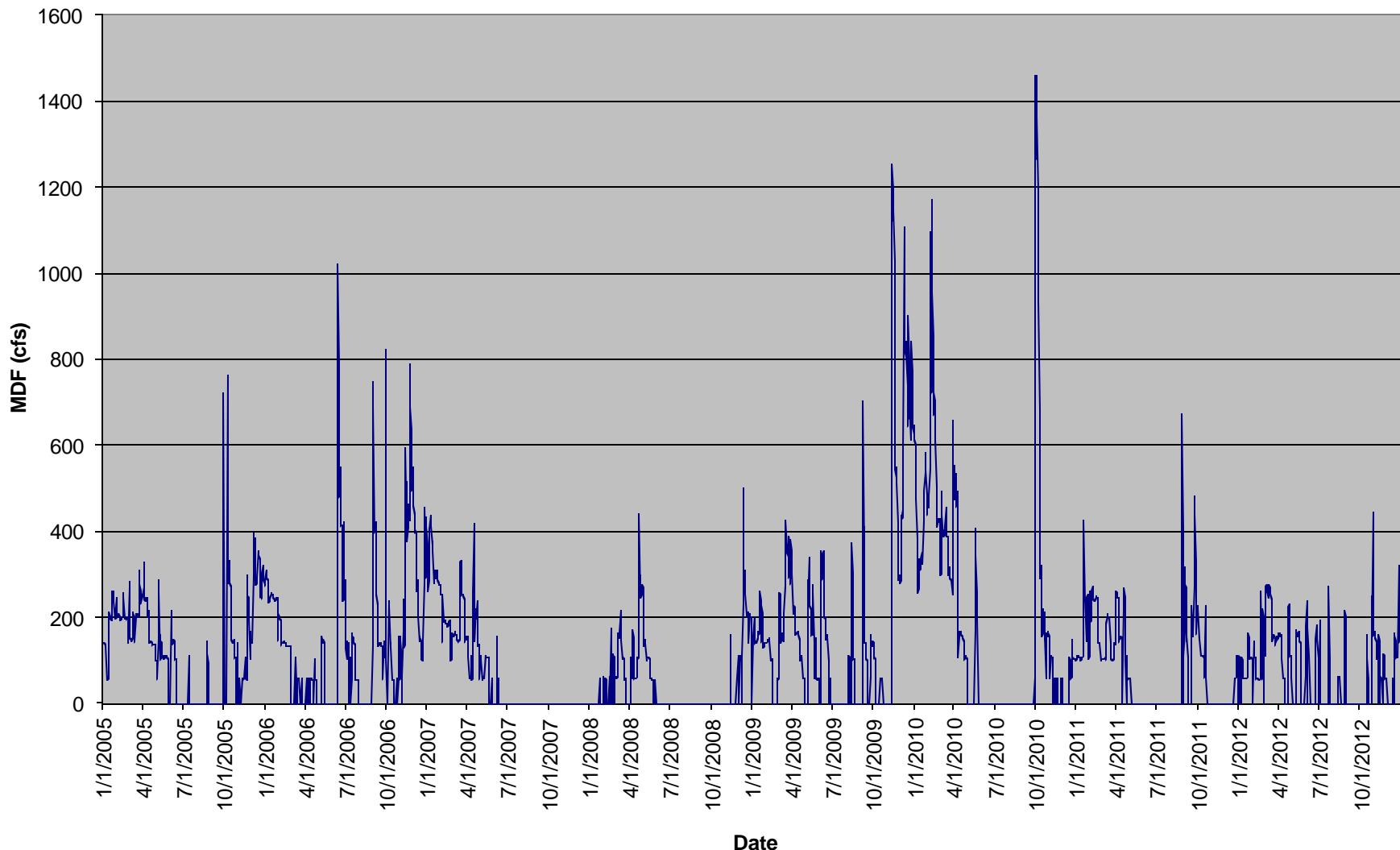


10 stations;
36 samples per station

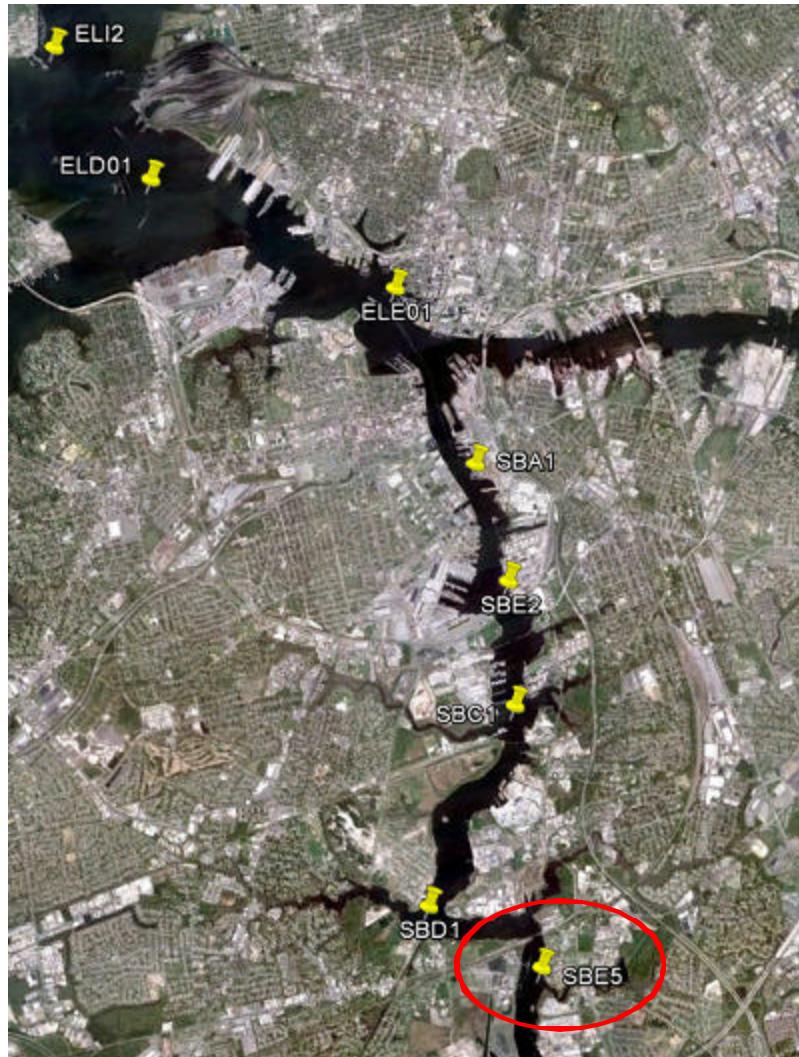
Deep Creek Spillway Flows



Deep Creek Spillway Flows



Elizabeth River SBE5



Elizabeth River SBE5

Correlation Matrix, Monthly ensemble average, 3 periods , 12 data points (12 months) per period per parameter

	time periods	Pearson Correlation Coefficients, N = 12																
		logChl-a	FLOW	WTEMP	SALINITY	TP	SECCHI	TSS	TN									
logChl-a	1991~2010	1																
	1991~2000		1															
	2001~2010			1														
FLOW	1991~2010		0.18627	1														
	1991~2000			0.26387	1													
	2001~2010				-0.0111	1												
WTEMP	1991~2010				0.18546	-0.81596	1											
	1991~2000					0.2734	-0.66566	1										
	2001~2010						0.11884	-0.88641	1									
SALINITY	1991~2010					-0.20876	-0.85505	0.71142	1									
	1991~2000						-0.48545	-0.83574	0.57242	1								
	2001~2010							0.20259	-0.80618	0.75526	1							
TP	1991~2010						0.01036	-0.57458	0.80485	0.72816	1							
	1991~2000							0.11918	-0.30033	0.78409	0.47552	1						
	2001~2010								0.00685	-0.75676	0.78425	0.80618	1					
SECCHI	1991~2010							-0.78515	-0.21526	-0.24496	0.39386	-0.08288	1					
	1991~2000								-0.77809	-0.26075	-0.37049	0.49975	-0.3095	1				
	2001~2010									-0.74397	-0.24999	0.01909	0.24704	0.21499	1			
TSS	1991~2010								0.59962	-0.28406	0.63881	0.07565	0.35805	-0.6804	1			
	1991~2000									0.59496	-0.36738	0.7326	0.10444	0.45694	-0.62301	1		
	2001~2010										0.57358	-0.14777	0.35304	-0.02696	0.12413	-0.63677	1	
TN	1991~2010									-0.34061	0.58446	-0.75405	-0.40654	-0.41459	0.28608	-0.45156	1	
	1991~2000										-0.06085	0.7224	-0.78315	-0.56233	-0.39648	0.05546	-0.51714	1
	2001~2010											-0.51511	0.40787	-0.3456	-0.3506	-0.06905	0.26985	-0.27604

Summary of regression analysis between Chl-a and each water quality parameter

Raw Data

Monthly (ensemble average, 1991~2010)

Statistical Model:

$$\log(CHLA) = a + bx + \epsilon$$

(Linear regression, log transform on Chl-a)

Parameters	R ²	samples
WTEMP	0.05	221
SALINITY	0.11	221
SECCHI	0.05	221
TSS	0.08	227
TP	<0.05	227
TN	<0.05	223
FLOW	<0.05	232

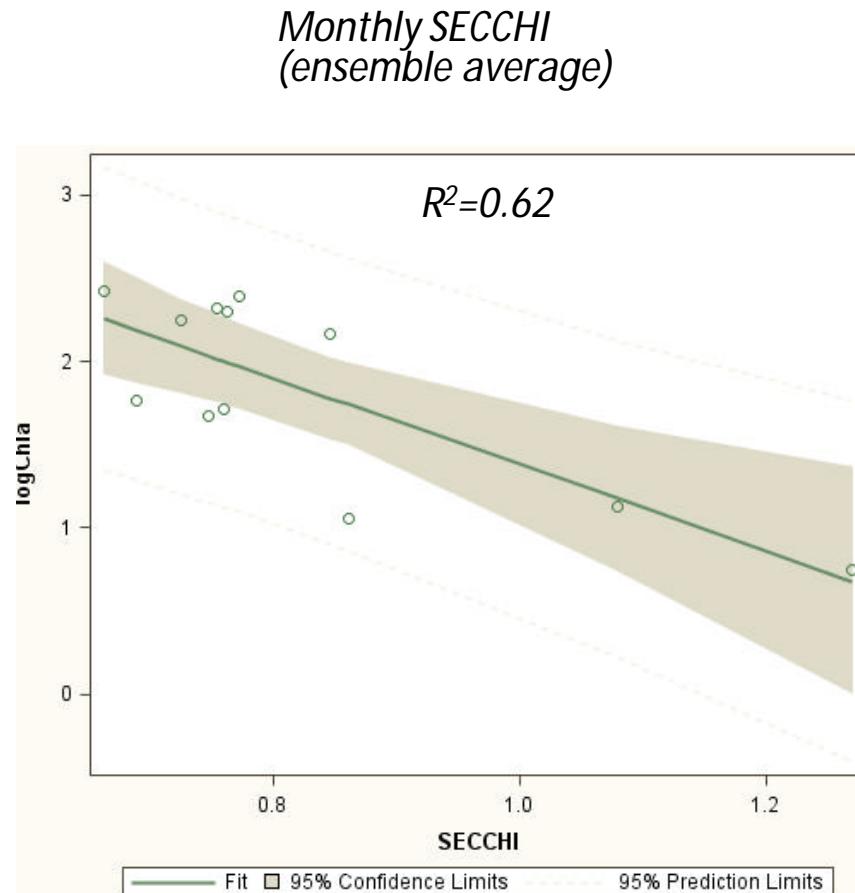
12 samples per parameter

Parameters	R ²
WTEMP	<0.05
SALINITY	<0.05
SECCHI	0.62
TSS	0.36
TP	<0.05
TN	0.12
FLOW	<0.05

12 samples per parameter

Elizabeth River SBE5

Selective regression plots between Chl-a and each water quality parameter
1991~2010 , (only those with high correlation)



Elizabeth River SBE5 Multiple regression (On sampling dates)

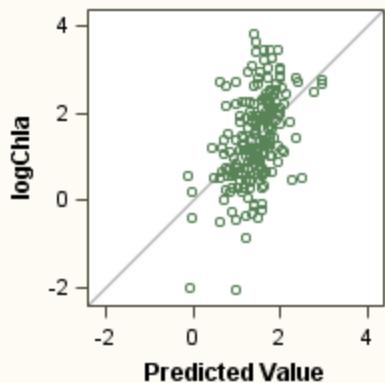
Multiple Regression
Statistical Model:

$$\log(CHLA) = a + b_i x_i + \epsilon$$

(1991~2010)

$R^2=0.22$

203 samples per parameter

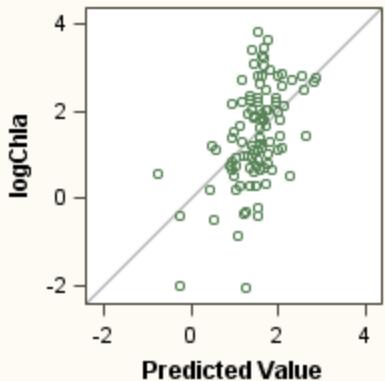


Variable	DF	Parameter Estimates				
		Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	0.80682	0.59435	1.36	0.1762	0
FLOW	1	-0.00121	0.000719	-1.68	0.0955	-0.18063
WTEMP	1	0.00075	0.01203	0.06	0.9503	0.00544
SALINITY	1	0.07212	0.02767	2.61	0.0099	0.32194
TP	1	-1.78006	2.38241	-0.75	0.4559	-0.05944
SECCHI	1	-0.61885	0.20504	-3.02	0.0029	-0.25117
TSS	1	0.03195	0.00989	3.23	0.0014	0.21258
TN	1	-0.063	0.21967	-0.29	0.7746	-0.02257

(1991~2000)

$R^2=0.25$

97 samples per parameter

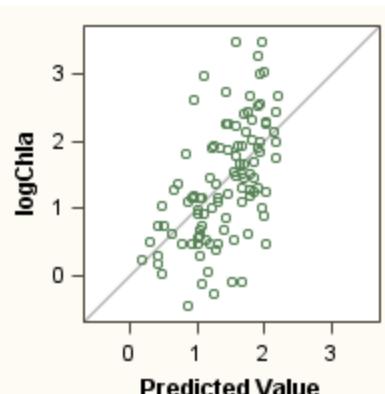


Variable	DF	Parameter Estimates				
		Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	1.96806	1.11867	1.76	0.082	0
FLOW	1	-0.00279	0.00104	-2.68	0.0089	-0.41805
WTEMP	1	-0.00495	0.01994	-0.25	0.8046	-0.03087
SALINITY	1	0.01543	0.04282	0.36	0.7194	0.06042
TP	1	-0.57492	3.92727	-0.15	0.8839	-0.01497
SECCHI	1	-0.73608	0.33285	-2.21	0.0296	-0.26377
TSS	1	0.0299	0.01356	2.2	0.03	0.21601
TN	1	0.0061	0.49658	0.01	0.9902	0.00162

(2001~2010)

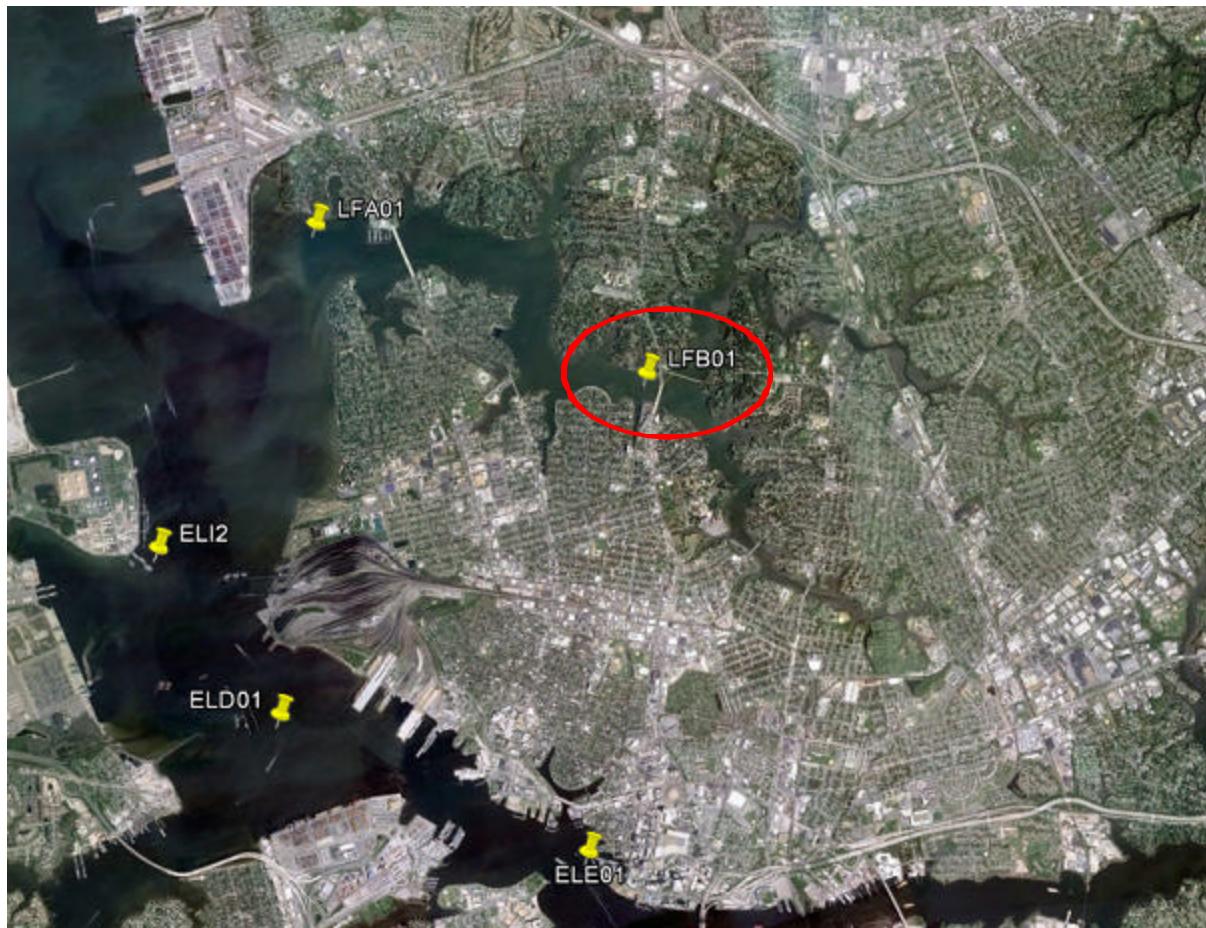
$R^2=0.31$

106 samples per parameter



Variable	DF	Parameter Estimates				
		Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	-1.27318	0.73656	-1.73	0.087	0
FLOW	1	0.0019	0.00108	1.77	0.0804	0.28291
WTEMP	1	0.01924	0.01468	1.31	0.1932	0.16462
SALINITY	1	0.18239	0.03817	4.78	<.0001	0.93047
TP	1	-6.43103	2.94713	-2.18	0.0315	-0.27413
SECCHI	1	-0.73553	0.24756	-2.97	0.0037	-0.34702
TSS	1	0.01636	0.01957	0.84	0.4053	0.07636
TN	1	-0.004	0.21555	-0.02	0.9852	-0.00186

Lafayette River LFB01



Lafayette River LFB01

Correlation Matrix, Monthly ensemble average, 1998~2010, 12 data points (12 months) per period per parameter

Pearson Correlation Coefficients, N = 12							
	Prob > r under H0: Rho=0						
	logChl-a	WTEMP	SALINITY	TP	SECCHI	TSS	TN
logChl-a	1	0.6935	0.57692	0.89789	-0.60406	0.68286	0.80712
WTEMP	0.6935	1	0.51172	0.78589	-0.8107	0.91568	0.81177
SALINITY	0.57692	0.51172	1	0.74312	-0.06453	0.29869	0.66081
TP	0.89789	0.78589	0.74312	1	-0.62065	0.74259	0.95829
SECCHI	-0.60406	-0.8107	-0.06453	-0.62065	1	-0.92357	-0.66064
TSS	0.68286	0.91568	0.29869	0.74259	-0.92357	1	0.74368
TN	0.80712	0.81177	0.66081	0.95829	-0.66064	0.74368	1

Lafayette River LFB01

Summary of regression analysis between Chl-a and each water quality parameter

Raw Data

Monthly (ensemble average, 1998~2010)

Statistical Model:

$$\log(CHLA) = a + bx + e$$

(Linear regression, log transform on Chl-a)

Parameters	R ²	Sample
WTEMP	0.13	143
SALINITY	<0.05	142
SECCHI	0.28	143
TSS	0.17	142
TP	0.34	138
TN	0.39	137

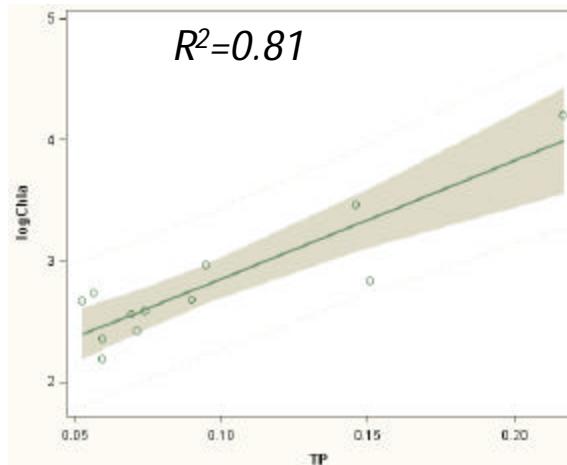
Parameters	R ²
WTEMP	0.48
SALINITY	0.33
SECCHI	0.36
TSS	0.47
TP	0.81
TN	0.65

12 samples per parameter

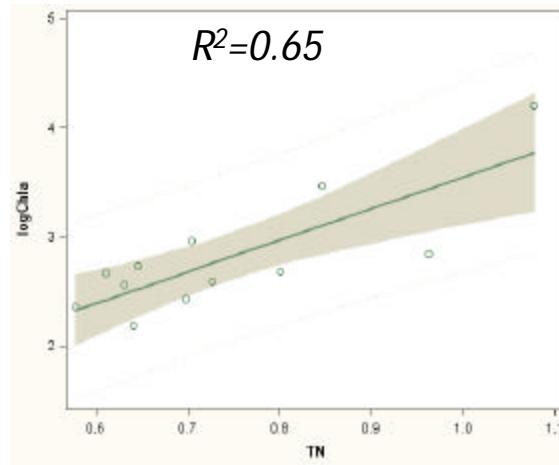
Lafayette River LFB01

Selective regression plots between Chl-a and each water quality parameter
1998~2010 , (only those with high correlation)

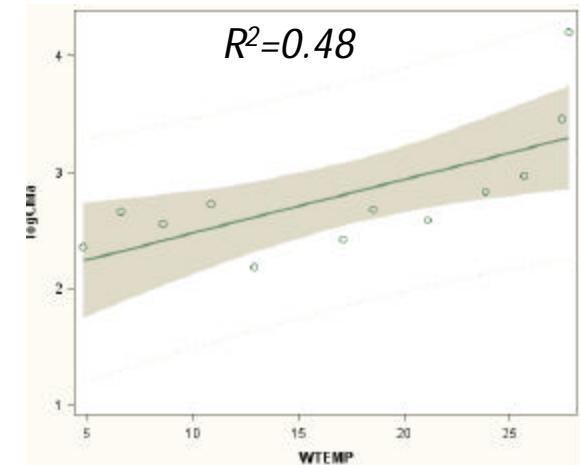
Monthly TP
(ensemble average)



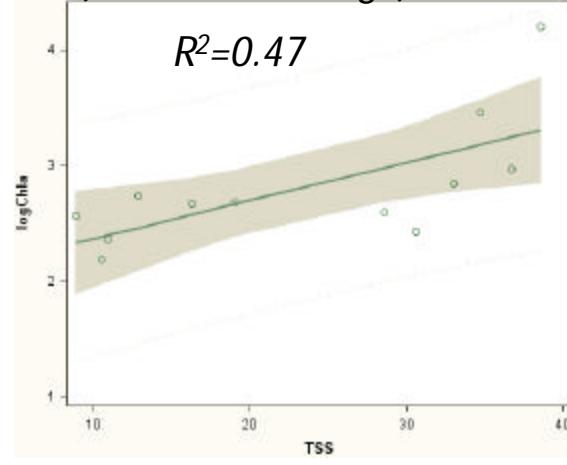
Monthly TN
(ensemble average)



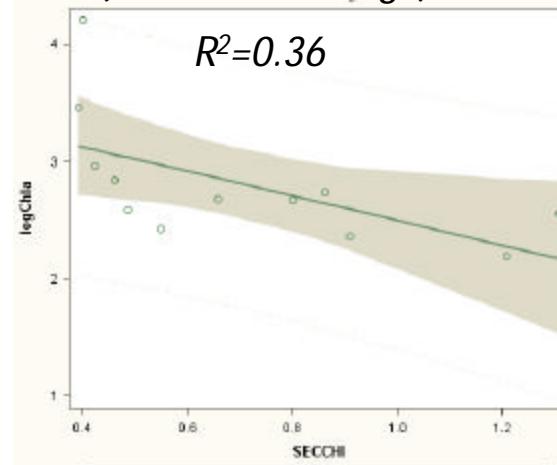
Monthly WTEMP
(ensemble average)



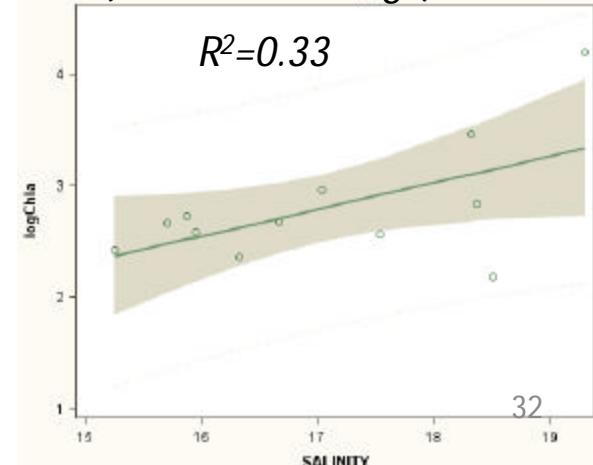
Monthly TSS
(ensemble average)



Monthly SECCHI
(ensemble average)



Monthly SALINITY
(ensemble average)



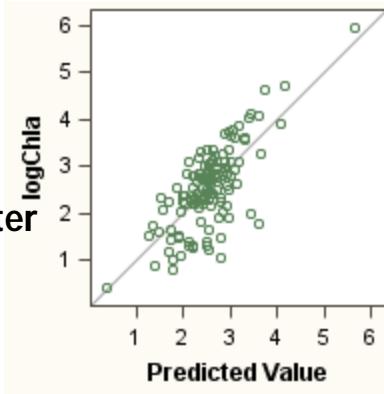
Lafayette River LFB01 Multiple regression (On sampling dates)

Multiple Regression
Statistical Model:

$$\log(CHLA) = a + b_i x_i + \epsilon$$

(1998~2010)

131 samples per parameter



Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	2.57508	0.44232	5.82	<.0001	0
WTEMP	1	-0.02037	0.00972	-2.1	0.0382	-0.19788
SALINITY	1	-0.00028	0.0172	-0.02	0.9869	-0.00125
TP	1	4.14604	1.31161	3.16	0.002	0.31673
SECCHI	1	-1.0461	0.22658	-4.62	<.0001	-0.46325
TSS	1	-0.007	0.00661	-1.06	0.2919	-0.10986
TN	1	1.09309	0.29024	3.77	0.0003	0.34974

Lafayette River LFB01

PCA (identifying the patterns among water quality parameters) On sampling dates (1998~2010)

Rotated Factor Pattern			
	Factor1		Factor2
TP	77*		26
TN	70*		-33
SECCHI	-82*		17
TSS	81*		3
WTEMP	82*		29
SALINITY	4		97*

Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.4 are flagged by an '*'.

2 main factors explaining 72% of the variance

Factor 1: Turbidity, temperature and nutrient dominated

Factor 2: salinity dominated

138 samples per parameter

Pearson Correlation Coefficients			
Prob > r under H0: Rho=0			
Number of Observations			
logChl-a	1	0.61896	-0.18298
		<.0001	0.0364
	144	131	131
Factor1	0.61896	1	0
	<.0001		1
	131	138	138
Factor2	-0.18298	0	1
	0.0364	1	
	131	138	138

Factor 1 has positive correlation with Chl-a: high water temperature, high nutrient concentrations and high turbidity condition leads to high Chl-a concentration. The R^2 is not very high though.

Factor 2's correlation is weak

Summary:

- The statistical analysis identified different behavior in different portions of the river system. Some clusters predict Chl-a with good scores using local water quality measurements, however other clusters have little skill in predicting Chl-a based on local water quality variables selected.
- Groups with good skill are the James above the TF5.5 and the upper Lafayette River both of which are relatively close to the loading source.
- The flow has a confounding role on flushing the Chl-a and delivering the nutrient load. For tidal fresh, the flushing dominates, leading to the inverse relationship between Chl-a and flow. For mesohaline and polyhaline, the delivery of nutrient dominates and Chl-a is positively related to flow.

- In the Upstream of Lafayette, the multi-regression R² are moderate (0.54) with TN (0.35), TP (0.31), and Secchi (-0.46) dominate, indicating the excess of nutrient inputs from the main metropolitan area.
- This analysis suggests that Chl-a in the downstream portion of the James and the Elizabeth River could be influenced by the exchange with Chesapeake Bay water quality condition.